

Fetch Take home assessment Part 3

Hey <Name>,

Hope you're doing well!

I want to share some key data quality issues and insights from our initial analysis, along with a few questions that require your input to ensure accuracy and completeness. Your feedback will be invaluable in refining our approach and driving better insights.

Data Quality Issues & Questions: We have identified a few data inconsistencies that may impact analysis accuracy. A detailed document is attached, but here are the key highlights:

- **Duplicate Entries:** The Transactions and Products datasets contain duplicate records, which we have de-duplicated/removed in our analysis. Could you confirm if these are data entry errors or if there is a preferred method for handling them?
- **Null Values:** Several columns across datasets have missing values. Do these nulls represent data that was not collected, or are they expected and non-applicable in certain cases?
- **Inconsistent Data Types:**
 - In the Transactions dataset, the FINAL_QUANTITY field contains a text value ('zero') instead of a numeric 0. Are there data validation rules in place to ensure type consistency?
 - In the Products dataset, a single BARCODE is associated with multiple categories/manufacturers, causing duplication when merging with transaction data. Is this expected behavior (e.g., brands switching manufacturers), and if so, how do we identify the latest manufacturer?

Key Data observations: Here are some early trends and insights we uncovered

- **User Behavior (21+ Age Group):** Brands such as Dove, Hershey's, Coca-Cola, Nerds Candy, and Great Value have the highest number of receipts scanned. Are there promotional incentives driving this behavior?
- **Health & Wellness Sales:**
 - ~99% of sales in this category lack birth date data, limiting our ability to segment sales by age groups. Would it be possible to improve data capture here?
 - Despite missing data, our analysis shows that Baby Boomers and earlier generations drive the majority of sales in this category, followed by Millennials and Gen X, while Gen Z has no recorded sales.
- **Dips & Salsa Category:** TOSTITOS leads across total sales, quantity purchased, receipts scanned, and store presence.

Next Step and Feedback request:

Your insights on these data issues and trends would be extremely helpful in ensuring our analysis is both accurate and actionable. We'd be happy to walk through these points in more detail, please let us know a convenient time for a discussion.

Looking forward to hearing back from you.

Thanks & Regards,

Raghav Goel
Data Analyst | Fetch

Submitted by Raghav Goel